

ERIC Document Resume		See Instruction Sheet		DATE OF RESUME 6/11/65		DATE RECEIVED	
A. PUBLICATION DATA							
1. ERIC NUMBER ED000761		2. ORIGINATOR CONTROL NO(S). 8205		3. IDENTIFICATION NO.		DO NOT WRITE IN SPACE BELOW	
4. AUTHOR(S): Becker, Samuel L., Murray, James N. Jr., and Bechtoldt, Harold P.				MICROFILM COPY AVAILABLE YES ☒ NO ☐			
5. TITLE OF DOCUMENT ABSTRACTED: Teaching by the Discussion Method: An Experiment With Television Discussions, Small Group Discussion, Large Group Discussion, and the Lecture Method				COPYRIGHT GRANTED YES ☐ NO ☒			
6. PUBLICATION TITLE: N.A.				COPYRIGHT RELEASE FILED YES ☐ NO ☐			
7. EDITOR(S): N.A.		8. PLACE OF ORIGIN ORG. ST. Univ. of Iowa CITY, STATE Iowa City, Iowa		TYPE OF REPORT PROPOSAL ☐ IN-PROGRESS ☐ FINAL ☒ FOLLOW-UP ☐			
9. PUBLISHER OR SOURCE Univ. of Iowa		TELE. EXT. INST/LCC CODE		OTHER ERIC REFERENCES:			
10. VOL. N.A.	11. PAGES N.A.	12. MO./YF. 58	13. TOTAL PP. 31	ERIC SATELLITE CODE			
B. INDEXING DATA							
14. SUBJECT AREA				15. TYPE OF ENDEAVOR experimental study			
16. INDEXING TERMS (6-12)		teachers		achievement			
experimental studies		attitudes		audio-visual aids			
teaching methods		ability		class size		teaching aids	
social studies		students		educational TV			
17. ABSTRACT (In With the aid of a grant from the Fund for the Advancement of Education, the University of Iowa experimented with various methods of teaching a course in American Government during the academic year 1956-57. The purpose of the study was to determine the answers to seven questions: 1. Can knowledge of and certain attitudes toward American Government be taught better by the the discussion or by the lecture method? 2. Does method of instruction differentially affect the acquisition of knowledge and attitudes of students of different levels of academic ability? 3. What methods of instruction are preferred by students? 4. What methods are preferred by instructors? 5. Can a class with an enrollment of 60-80 be taught by the discussion method and achieve the same results in course examinations as a class of 20-25 taught by the same method? 6. Can an instructor using closed-circuit television promote more discussion and secure better results in course examinations that he can with the same number of students in a conventional classroom situation? 7. Will performances of students who can see and hear a television discussion, but who cannot participate orally in it, be differently affected as compared to those who can participate? To answer these questions five methods of teaching were compared: Television discussion, television observation, small-group discussion, large-group discussion, and lecture. Within limitations described in the report, findings supported the hypothesis that the five methods of teaching did not differ significantly in their effect on acquisition of knowledge and attitude change. It was found, however, that the students preferred a course taught at least partially by the discussion method.							
ERIC No. 18. SOURCE OF ABSTRACT							
C. FUNDING DATA							
19. SUPPORTING ORGANIZATION				20. DIVISION OR DEPT.		21. CONTRACT NO.	
				22. TOTAL COST OF PROJECT		PROJECT NO.	
				23. DURATION OR PROJECT		U.S. Department of Health, Education, and Welfare Office of Education	

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education

This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.



*Teaching
by the
Discussion
Method*

STATE UNIVERSITY OF IOWA

ED 000761

*Teaching by the
Discussion Method*

• • •

An Experiment with Television Discussion, Small Group
Discussion, Large Group Discussion, and the Lecture Method.

SAMUEL L. BECKER

Director of the Division of Television, Radio, Film

JAMES N. MURRAY, JR.

Assistant Professor of Political Science

HAROLD P. BECHTOLDT

Associate Professor of Psychology

Supported in part by a grant from the
Fund for the Advancement of Education

• • •

STATE UNIVERSITY OF IOWA • 1958

Introduction

The State University of Iowa, with the aid of a grant from the Fund for the Advancement of Education, experimented during the academic year 1956-57 with various methods of teaching American Government. Prior to the experiment proper, "trial runs" had been made in both American and comparative Government courses, as well as in courses in Literature. These aided in pointing up and reducing the problems involved in these methods of instruction, especially in the use of closed-circuit television for discussion. They also made it clear that this use of television was feasible. The purpose of the 1956-57 experiment, here reported, was to evaluate more precisely the achievement in the usual academic activities as a function of the method of instruction and to determine the effect of each method on certain attitudes of students and staff.

The problem studied in this experiment was selected for several reasons. First, there seems to be a dearth of information available about the effectiveness of various teaching methods on the college level. One question which often arises is whether the discussion method of teaching does result in skills and attitudes which are superior to those which result from other methods, such as the lecture. Second, it is a well-known fact that no single student is able to participate very often in a classroom discussion. In other words, much of the discussion experience is actually vicarious. Therefore, does it make any difference whether this experience is gained by actual presence in the classroom or can it be obtained just as well via television? Third, in view of the rising enrollment problem it seems doubtful whether staff and facilities can be found to do a great deal of teaching by the small group discussion method as customarily conceived unless some technique such as closed-circuit television can come to the rescue. In brief, this research program was undertaken not simply to study closed-circuit television but rather to utilize television in the study of some educational problems which have confronted higher education for a considerable period of time and which appear likely to become more acute in the years immediately ahead.

This report is confined to the results of the 1956-57 experiment, actually one experiment made during the first semester and replicated during the second semester. Part I of the report summarizes the major results. Part II outlines the nature of the experiment. Part III indicates the measured results in greater detail. Part IV points out the limitations of the study. Part V cites the conclusions.

PART I

Summary of Results

These were the major results of the 1956-57 experiment:

A) Achievement in the American Government course, as measured by the mid-term and final examinations of this course, was improved by all methods of instruction. It made no difference whether the method was television discussion or observation only of the televised course, whether it was small group discussion, large group discussion, or lecture.

B) Students of high and low ability were not differentially affected by the method of instruction. The results, however, were consistent with the results of most other educational research in showing the importance of general academic skill or ability. If the method of instruction made the slightest difference, this difference was dwarfed by the general academic ability factor.

C) All methods of instruction were apparently successful in affecting attitudes toward the concepts of "liberal democracy." The changes were in a direction previously identified by the course instructors as favorable. Again, there were no significant differences between methods of instruction.

D) Students in this course preferred the course taught by the discussion method or a combination of lecture and discussion, rather than lecture only.

E) Students who had experienced a course taught by the discussion method on television were more favorable toward the use of television for instructional purposes than were students who had not had this experience.

F) The students' ratings or self-evaluations of the amount they learned as well as the degree to which they were stimulated by the course were positively related, in the television sections, to the students' opportunities to participate.

These results were consistent with the results of other experimentation carried out at the State University of Iowa during the 1954-55, 1955-56, and 1956-57 academic years.

It appeared as feasible to conduct a class by the discussion method with 75 to 80 students, either in a normal classroom situation or by means of closed-circuit television, as with 18 to 25 students. Closed-circuit television, as used in this experiment, neither deterred nor aided discussion with 75 to 80 students. The only advantage of discussion over lecture appeared to be that students tended to prefer the discussion method. The instructors, on the other hand, preferred relatively more lecturing.

PART II

The Experiment

The course. American Government, a three hour course meeting three days a week for fifty minutes at a time, is considered the basic political science course at the State University of Iowa. It is required of majors in the department as well as of students in certain other areas of study, notably teacher training. It is open to Freshmen, although ordinarily as many Sophomores as Freshmen are enrolled, and there are a number of Juniors and Seniors as well.

Traditionally the course has been taught by what is called the lecture-discussion method in sections varying in size from approximately fifty to one hundred and twenty-five students. Section size has depended on course enrollment and the availability of instructors. The extent of class discussion has tended to vary inversely with the number of students and, to a lesser degree, directly with the personal predilections of the instructor. One further point should be noted here; other than on the most broadly conceived topics for discussion, there exists at Iowa—and in the profession at large—no consensus as to the content of American Government courses. While most instructors would agree that such subjects as drafting the constitution, federalism, civil rights, political parties, pressure groups, and each of the three main branches of our national government ought to be “covered,” there are varied possibilities of dealing with each, depending on the interests, competence, values, and methodological orientation of the instructor. This fact has certain implications for this experiment and for many experiments concerned with teaching methodology.

Purpose of the study. After having demonstrated in the trial runs that the course could be taught by the discussion method and that the use of closed-circuit television for courses taught by the discussion method was physically possible, the experiment was designed to answer seven questions in terms of the criteria mentioned above and described in detail later.

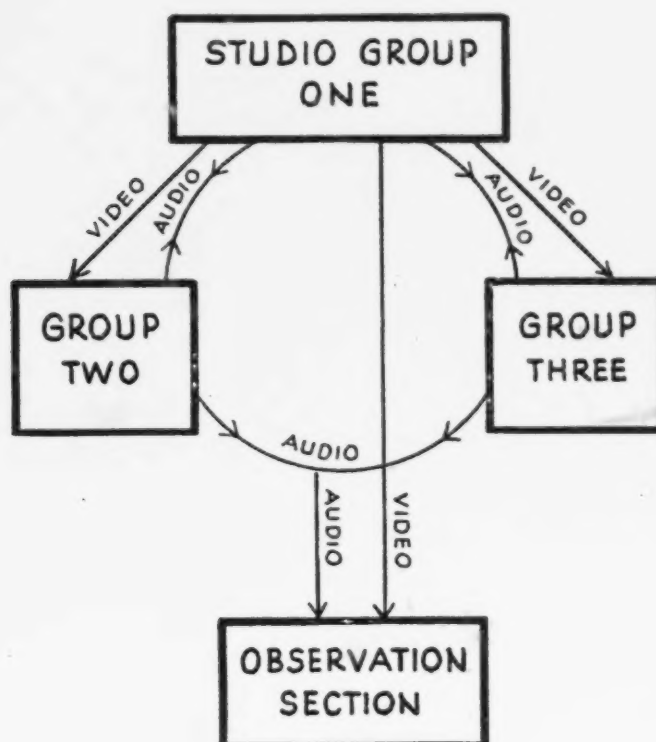
1. Can knowledge of and certain attitudes toward American Government be taught better by the discussion method or by the lecture method?
2. Does the method of instruction differentially affect the acquisition of knowledge and attitudes of students of different levels of academic ability?
3. What methods of instruction are preferred by students in the American Government course?

4. What methods of instruction are preferred by instructors of the American Government course?
5. Can a class with an enrollment of sixty to eighty be taught by the discussion method and achieve the same results in course examinations as a class of twenty to twenty-five taught by this method?
6. Can an instructor, using closed-circuit television, promote more discussion and secure better results in course examinations than he can with the same number of students in a conventional classroom situation?
7. Will the performances of students who can see and hear a television discussion, but who cannot participate orally in it, be differently affected as compared to those of students who can participate?

Treatment groups. In order to answer the above questions, five types of sections were compared.

1. *Television discussion.* In this section a comparatively large number of students (eighty-six the first semester, sixty-six the second semester) was divided into three groups. Group One consisted of fourteen students who met with the instructor in the television studio. The members of this group, along with the instructor, were "on camera," both their images and voices being relayed, via closed circuit television, to monitors in other rooms. Groups Two and Three were in two of these other rooms. There they could see and hear what was going on in Group One and could communicate with Group One and with each other. Group One, however, could see neither of these Groups. Thus discussion could be carried on among the three Groups and Groups Two and Three could see Group One in action. This plan can be seen graphically in Figure 1. Normally, the burden of responsibility for discussion fell on members of Group One, but the instructor often called on students in the other two groups and, of course, welcomed volunteers. A graduate assistant, who served as proctor, was assigned to each of the other rooms. Students in Groups Two or Three indicated a desire to enter the discussion by raising their hands. The proctor would indicate this to the instructor by pressing a button which would flash either a red or white light (depending on which group wanted to be heard) and the instructor would call upon that room. The student would then identify himself and proceed to talk. Students were rotated among the three groups approximately every two weeks. All students spent approximately an equal amount of time in each group.

2. *Television observation.* One section of students (twenty-seven the first semester, twenty-four the second semester) was placed in a room in which Group One could be seen and heard via television monitors and the discussion among Groups One, Two, and Three could be heard. Members of this group could *not*, however, participate in the discussion. That is to say, they could not be seen or heard by any of the other television groups.



3. *Small group discussion.* Most proponents of the discussion method, and opponents for that matter, agree intuitively that small group discussion is more effective than large group discussion. Hence it was thought useful to have this type of section as a basis for comparison with other types using the discussion method. Three such sections were employed during the first semester, two during the second semester. Enrollments in these sections varied from eighteen to twenty-nine.

4. *Large group discussion.* This section was designed to show how a group roughly as large as that in the television discussion section, but not divided into separate groups, would compare with the television section and the small discussion sections. There were seventy-five students in this section each semester.
5. *Lecture.* As a measure of comparison for the discussion method of teaching in any of the above forms, two large sections (one hundred and twelve to one hundred and thirty-two students each) were taught each semester by the lecture method. In each case the instructor allowed interruptions from the class for questions or comments, but in general the size of each section seemed to preclude any systematic discussion of topics except by the instructor.

Four additional points need to be noted with reference to these sections.

In the first place, it was impossible to utilize strictly random sampling procedures except in the assignment of students to one of the television section groups. The students were randomly assigned from one large group to either the television observation or television discussion sections. Assignment of others depended on the hour at which each chose to take the course. The assumption was made that students who registered for the course at different times of day did not initially differ significantly on any of the variables which would affect the criterion measures.

Secondly, four teachers (one instructor and three assistant professors) were involved in the experiment. One taught all of the small group discussion sections and one of the large lecture sections during the second semester. One taught a lecture section the first semester. One taught a lecture section each of the two semesters. One taught the television sections and the large group discussion sections. It obviously would have been better to have more than one instructor involved in each type of section but scheduling difficulties precluded this.

Thirdly, obviously not all possible types of sections were included in the experiment. Most notable by its absence was the "traditional" lecture-discussion section of fifty to sixty students.

Finally, an extensive effort to promote discussion and to provide a common core of topics covered in the course was made in the form of a rather elaborate syllabus drawn up by members of the political science department.¹ This included topical (almost daily) assignments, objectives, and (most important) questions covering the material for which the student was to be held responsible and which served as the basis for discussion on any given day in the discussion sections. This, in the opin-

ion of the instructors concerned, helped to promote discussion to a marked degree.



Evaluation criteria. Four criteria were utilized for a comparison of the five groups.

1. Achievement, as measured by two mid-term and one final examination. Each mid-term contained forty multiple-choice items and an essay question. The first semester final examination included one essay question and eighty multiple-choice questions. The second semester final was composed of seventy multiple-choice questions and one essay question. Forty of these multiple-choice questions and the essay question were also given to each group at the beginning of each semester so that a measure of information gain could be used for the comparison. The course instructors agreed, prior to the testing of the data, that these achievement examinations were an adequate measure of achievement in this course.
2. Attitudes toward the so-called concepts of "liberal democracy," as revealed by a measure being developed by Assoc. Prof. Lane Davis of the political science department. The forty-eight item Likert-type questionnaire was administered at the beginning and end of each semester.

3. Opinions concerning the various methods of instruction and the perceived effects of each method. These opinions were assessed in a questionnaire completed by each student during the last two weeks of the first semester in which the experiment was conducted.

4. Attitudes of the discussion instructors toward these methods of instruction.

The three discussion methods were also compared on the basis of class participation.

The hypothesis tested was that the five methods of instruction did not differ significantly on any of the four criteria noted above.

PART III

The Results

Results on the achievement criterion.

The data secured in this study necessitate the retention of the hypothesis that these methods of instruction do not differ significantly on the criterion of achievement, as defined above. The evidence also supports the assumption that these groups initially did not differ significantly on any of the variables assumed a priori to affect the criterion measures.

The basic results can be seen in Tables I and II. The first column of each table shows the number of students enrolled in each section of the course during that semester. The second column shows the mean Entrance Composite Percentile Rank for the students in each section. The homogeneity of the sections on this criterion was tested each semester with an analysis of variance. These groups did not differ significantly on the Entrance Composite Percentile Rank which each student had received when he entered the University. The F-ratio was less than unity for the first semester data and was 1.07 for the second semester data.

Columns 3, 4, 5, and 6 of Table I show the mean score of each section during the first semester on each of the three achievement examinations plus the sum or total of the three. Table II shows this information for

TABLE I

Means for Entrance Composite Percentile Rank and Achievement Test Scores for 389 Cases Participating in the First Semester of the 1956-57 Course in American Government²

Group	1 No. of Cases	2 Entrance Composite	3 1st Ach. Test	4 2nd Ach. Test	5 3rd Ach. Test	6 Ach. Total
Lecture	105	51.86	39.85	35.02	59.08	133.97
Television						
Discussion	63	52.14	39.93	35.73	56.84	132.50
Television						
Observation	22	45.90	41.31	35.77	59.09	136.18
Lecture	89	51.23	40.34	35.50	58.88	134.74
Small Group						
Discussion	19	61.26	43.42	39.10	63.10	145.63
Small Group						
Discussion	16	55.62	40.50	35.37	59.06	134.93
Large Group						
Discussion	55	48.87	41.03	36.20	59.12	136.36
Small Group						
Discussion	20	51.95	40.35	37.55	59.00	136.90

TABLE II

Means for Entrance Composite Percentile Rank and Achievement Test Scores for 420 Cases Participating in the Second Semester of the 1956-57 Course in American Government

Group	1 No. of Cases	2 Entrance Composite	3 1st. Ach. Test	4 2nd Ach. Test	5 3rd Ach. Test	6 Ach. Test
Lecture	115	52.88	39.00	40.63	61.30	140.94
Television						
Discussion	59	53.96	39.81	40.67	62.13	142.62
Television						
Observation	24	50.54	40.70	42.70	61.30	146.38
Lecture	105	50.26	40.92	41.33	60.37	143.25
Small Group						
Discussion	28	52.64	39.75	41.35	60.96	142.07
Large Group						
Discussion	71	48.11	40.15	40.56	61.74	142.46
Small Group						
Discussion	18	66.29	41.61	42.60	62.17	146.38

the second semester. Tables III and IV summarize the analyses of variance on these data; no significant interaction between the various measures and the various sections was found; no significant differences among the sections on their achievement in these tests were indicated. The only significant F-ratio, as expected since a much higher score was possible on the third achievement examination, was between the various measures. The other F-ratios were less than unity both semesters.

TABLE III

Summary Table of Analysis of Variance between First Semester Groups on each Measure of Achievement: first and second mid-terms, objective items in the final which were in the pre-test, new objective items in the final, and essay questions in the final examination.³

Source of Variation	df	Sums of Squares	Mean Squares	F	Sig. at 5% Level of Conf.
Between-Subjects	388	32,047			
Groups (G)	7	571	81.57	<1	
error (b)	381	31,476	82.61		
Within-Subjects	1556	227,052			
Measures (M)	4	201,383	50,346	3,042	Sig.
MG	28	445	15.89		
error (w)	1524	25,224	16.55		
Total	1944	259,099			

TABLE IV
Summary Table of Analysis of Variance between Second Semester Groups on each Measure of Achievement: first and second mid-terms, and final examination.

Source of Variation	df	Sums of Squares	Mean Squares	F	Sig. at 5% Level of Conf.
Between-subjects	419	68,581.76			
Groups (G)	6	322.76	53.79	<1	
error (b)	413	68,259.00	165.27		
Within-Subjects	840	141,688.67			
Measures (M)	2	123,274.38	61,637.19	2,801.69	Sig.
MG	12	239.14	19.92	<1	
error (w)	826	18,175.15	22.00		
Total	1259	210,270.43			

Since some have claimed that students of varying ability are affected differentially by different methods of instruction, it was decided to use a "treatments by levels" analysis of variance in the evaluation of the data from the replication of the American Government study. Two separate analyses were made using the final examination score as the criterion variable. On one of these analyses, the Entrance Composite Percentile Rank was used as the control variable while the Pre-test examination score was used as the control for the other. On both, cases were randomly eliminated in order to meet the requirement for proportionality in this type of analysis.⁴ In order to have enough cases in each cell, the two small discussion sections were grouped for these analyses. Three levels were used, the top quartile, the bottom quartile, and the middle two quartiles. These analyses are summarized in Tables V and VI. It can be seen from these summaries that the hypothesis of no difference between the effectiveness of the various methods of instruction cannot be rejected for any of the levels. The only significant differences were found between levels, which was consistent with other findings of this study.

TABLE V
Summary Table of Analysis of Variance of Final Achievement Scores for Second Semester Groups and Entrance Composite Percentile Rank Levels.

Source of Variation	df	Sum of Squares	Mean Squares	F	Sig. at 5% Level of Conf.
Groups (G)	5	141.59	28.32	<1	
Levels (L)	2	8,804.09	4,402.04	105.62	Sig.
Cells	17	9,362.44			
Groups by Levels	10	416.76	41.68	<1	
Within-subgroups (w)	342	25,105.96	73.40		
Total	359	34,468.40			

TABLE VI
Summary Table of Analysis of Variance of Final Achievement Scores for
Second Semester Groups and Pre-test Score Levels.

Source of Variation	df	Sum of Squares	Mean Squares	F	Sig. at 5% Level of Conf.
Groups	5	646.72	129.34	1.64	Sig.
Levels	2	5,303.66	2,651.83	33.59	
Cells	17	6,418.78			
Groups by Levels	10	468.40	46.84	<1	
Within-subgroups (w)	326	25,737.22	78.95		
Total	343	32,156			

Table VII lists the means for the total and each part of the Pre-test and Post-test for the first semester. The results of the analysis of variance of these data as well as those from the second semester were consistent with the hypothesis of no initial difference between groups on the achievement pre-test total. Analyses of achievement gain likewise indicated no differences among groups. In all cases, the F-ratios were less than unity. Tests of the homogeneity of test-retest correlations, run for the first semester data, showed no significant differences among the groups. Average correlations of the Pre-test and Post-test were .43 for the 40 objective items, .41 for the essay items, and .53 for the total.

TABLE VII
Means for Entrance Composite Percentile Rank and Pre- and Post-Achievement Test Scores for 389 Cases Participating in the First Semester of the 1956-57 Course in American Government.

Group	1	2	3			6		
	No. of Cases	Entrance Composite	Objective	Essay	Total	Objective	Essay	Total
Lecture	105	51.86	16.22	5.50	21.73	27.04	13.80	40.85
Television								
Discussion	63	52.14	15.50	6.39	21.90	26.66	12.55	39.22
Television								
Observation	22	45.90	15.22	6.00	21.22	28.18	12.09	40.27
Lecture	89	51.23	15.15	6.61	21.67	26.97	13.40	40.38
Small Group								
Discussion	19	61.26	16.57	7.73	24.31	28.05	13.89	41.94
Small Group								
Discussion	16	55.62	16.25	5.87	22.12	26.81	13.12	39.93
Large Group								
Discussion	55	48.87	16.43	6.25	22.69	26.56	13.21	39.78
Small Group								
Discussion	20	51.95	15.80	5.95	21.75	27.20	13.25	40.45

In view of the small between-groups variance found for each of these analyses, a covariance analysis using entrance composite as a control variable was not considered necessary.

Intercorrelations among the Achievement Test scores and between the Total Achievement scores and the Entrance Composite Percentile Ranks were calculated for the first semester data. These were consistently high. The within-groups correlations for these measures, together with the average correlations, (computed as an average of weighted transformed correlations) are presented in Table VIII. All correlations between

TABLE VIII

Intercorrelations Among the Variables of Entrance Composite Percentile Rank and Achievement Test Scores.

Group	Degrees of Freedom	Correlation between the			
		Entrance and Total Ach.	1st and 2nd Ach.	1st and Final Ach.	2nd and Final Ach.
Lecture	102	.55	.66	.70	.70
Television Discussion	60	.54	.46	.54	.49
Television Observation	19	.44	.60	.68	.65
Lecture	86	.53	.57	.51	.57
Small Group Discussion	16	.61	.33	.60	.68
Small Group Discussion	13	.52	.72	.56	.45
Large Group Discussion	52	.48	.63	.78	.67
Small Group Discussion	17	.45	.58	.56	.58
Total	365	.53	.59	.63	.62
χ^2 (test for homogeneity of variance) None are significant		0.1873	6.0819	11.0136	6.1423

groups were homogeneous and could be adequately represented by the average value given in the Total row toward the bottom of the table. It is interesting to note that the correlation between the Entrance Composite Percentile Ranks and the Total Achievement scores was almost as high as the intercorrelations among the Achievement scores. The

Entrance Composite Percentile Ranks were also correlated with each of the Pre-test and Post-test scores and the individual Achievement test scores. These results are shown in Table IX. The high correlations of the Entrance Composite with each of these measures raises a serious question about the sources of variance associated with grades in the course. From the high correlations obtained, one could argue that all of these measures can be interpreted in terms of a single hypothesized variable defined as the general academic skill and interest of the students. In terms of this hypothesis, levels of student performance in American Government can be said to be associated more closely with differences between individual students in academic skills than with differences between these methods of instruction.

TABLE IX
Correlation of Entrance Composite Percentile Rank with each Achievement and Pre-Test Measure.

	Mean Correlation for all groups
a. Pre-Test Objective (40 items)	.42
b. Pre-Test Essay	.31 ⁵
c. Pre-Test Total	.45
d. 1st Achievement Test	.51
e. 2nd Achievement Test	.40
f. Post-Test Essay (Same as b)	.36
g. Post-Test old Objective (Same items as a)	.40
h. Post-Test Total (f plus g)	.47
i. Post-Test new Objective (30 new items)	.29
j. 3rd Achievement Total (h plus i)	.45
k. Course Achievement Total (d plus e plus h)	.53

Results on the attitude criterion.

Table X shows the means for each of the first semester groups on the pre- and post-attitude measures. An analysis of variance, summarized in Table XI, indicated that attitudes changed significantly in a "favorable" direction but that there were no significant differences between groups on this change. The fact that the interaction was not significant would be consistent with the hypothesis of homogeneity of the means on both the pre- and the post-attitude measures. The mean correlation of the Entrance Composite with each of these tests was .41 and .42.

TABLE X

Means for Pre- and Post-Attitude Test Scores for 389 Cases Participating in the First Semester of the 1956-57 Course in American Government.

Group	No. of Cases	Pre-Test Attitude	Post-Test Attitude
Lecture	105	33.01	33.81
Television Discussion	63	30.34	33.20
Television Observation	22	30.45	34.18
Lecture	89	30.75	33.51
Small Group Discussion	19	33.94	35.10
Small Group Discussion	16	32.81	33.87
Large Group Discussion	55	30.21	32.30
Small Group Discussion	20	28.90	33.80

TABLE XI

Summary Table of Analysis of Variance of First Semester Differences Between Pre-and Post-Attitude Scores.

Source of Variation	df	Sums of Squares	Mean Squares	F	Sig. at 5% Level of Conf.
Between-Subjects	388	31,918			
Groups (G)	7	665	95	1.16	
error (b)	381	31,253	82		
Within-Subjects	389	7,937			
Test (A)	1	915	915	51.49	Sig.
AG	7	251	35.86	2.02	
error (w)	381	6,771	17.77		
Total	777	39,855			

The second semester attitude data were analyzed in a different way. A levels analysis was made, using the pre-attitude score as the control factor and the post-attitude score as the criterion measure. Table XII summarizes the results. Again the significance of the differences between levels is to be expected. Differences between groups were also found to be significant, however. The highest mean on the post-attitude measure was obtained by one of the lecture groups; the lowest was ob-

tained by the television observation group. Table XIII shows the differences between each pair of means. It can be seen in this Table that Lecture Group 2 scored significantly higher on the post-attitude measure than any of the other groups. None of the other groups differed significantly. A comparison of these results with those from the first semester leads one to believe that the size of this difference on the replication was a chance effect. During the first semester, one of the small discussion groups and the television observation group had the highest mean scores on the post-attitude measure.

TABLE XII

Summary Table of Analysis of Variance of Second Semester Post-Attitudes Measures for Groups and Pre-Attitude Levels.

Source of Variation	df	Sums of Squares	Mean Squares	F	Sig. at 5% Level of Conf.
Groups	5	618.58	123.71	5.30	Sig.
Levels	2	3,924.72	1,962.36	84.22	Sig.
Cells	17	4,626.10			
Groups by Levels	10	82.80	8.28	<1	
Within-subgroups (w)	258	6,013.67	23.30		
Total	275	10,639.77			

TABLE XIII

Table of Differences between Group Means on the Post-Attitude Measure.

Means:			G ₂	G ₃	G ₄	G _{5,7}	G ₆
Lecture 1	M ₁ -35.60	G ₁	.55	2.23	2.56*	.56	.92
TV Disc.	M ₂ -35.05	G ₂		1.68	3.11*	.01	.37
TV Obs.	M ₃ -33.37	G ₃			4.79*	1.67	1.31
Lecture 2	M ₄ -38.16	G ₄				3.12*	3.48*
Small Group Disc.	M _{5,7} -35.04	G _{5,7}					.36
Large Group Disc.	M ₆ -34.68	G ₆					

* Significant.

The pre-attitude scores of the first semester were also correlated with the scores from the final examination. The mean correlation was .31, but the correlations for individual groups ranged from .06 to .69. the χ^2 test for homogeneity of the correlations, however, just missed being significant at the 5 per cent level of confidence, and so the hypothesis of homogeneity must be retained.

Results on the opinion criterion.

Student opinions concerning the various methods and the effects of each method were secured only from the students registered for the course during the first semester. These opinions were consistent with data obtained from other experiments at the State University of Iowa during the period from 1955 to 1957.

Table XIV shows the opinions of each group toward the alternative methods of teaching this course. The percentages are based on the weighted responses. Zero points were given for a last choice, one point for next to the last, and so forth. Two results seem important to note. One was the preference of students for some form of discussion, either complete discussion in a regular classroom or lecture with one-quarter of the time for discussion. The other important result was the tendency of those students who had had some familiarity with the television discussion method to prefer this method. This was even true of the television observation group.

TABLE XIV

	TV Disc.	TV Obs.	Large Class Disc.	Lecture	Small Class Disc.
If American Govt. was being offered next year in a number of different ways and a friend of yours wanted to take it, which of the following would you advise him to take? (No. in the order of your choice.)	n=80	n=27	n=65	n = 249 n _A =117 n _B =132	n = 65 n _A =22 n _B =20 n _C =23
Section taught predominantly by straight lecture in a regular classroom.	13.2%	16.5%	22.3%	20.0%	20.7%
Section taught predominantly by the discussion method in a regular classroom.	26.5%	25.1%	26.0%	23.0%	30.7%
Section taught predominantly by straight lecture on closed-circuit TV.	10.0	9.4	8.7	8.3	6.2

	TV Disc	TV Obs.	Large Class Disc.	Lecture	Small Class Disc.
Section taught predomi- nantly by the discus- sion method with closed- circuit television and a sound system which enables students view- ing TV sets to enter the discussion.	29.7	25.8	10.9	15.2	12.1
Section taught by the lecture method primarily but with one-quarter of the time for questions and discussion	20.3	23.2	30.0	32.0	29.8
Makes no difference	0.3		2.1	1.5	.5

Differences between groups on the estimated relative preparation required for the course taught by each method are shown in Tables XV and XVI. The differences between groups in Table XV are statistically significant at the 5 per cent level of confidence. Those in Table XVI are not. ($\chi^2 = 26.21$ and 9.23 respectively. $df. = 8$. In all analyses, "no response" cases were discarded since $n < 5$ in these cells.)

TABLE XV

	TV Disc.	TV Obs.	Large Class Disc.	Lecture	Small Class Disc.
Do you feel that it is neces- sary to do more preparation for this course than for com- parable discussion courses you are taking or less?					
More for comparable discussion courses	21.3%	33.3%	13.8%	31.3%	7.7%
More for this course	22.5	11.1	32.3	24.5	29.2
About the same	56.2	51.9	53.9	42.6	60.0
No response		3.7		1.6	3.1

TABLE XVI

	TV Disc.	TV Obs.	Large Class Disc.	Lecture	Small Class Disc.
Do you feel that it is necessary to do more preparation for this course than for comparable <i>lecture</i> courses you are taking or less?					
More for comparable lecture courses	13.8%	25.9%	12.3%	16.5%	18.5%
More for this course	43.8	18.5	35.4	31.3	36.9
About the same	41.2	55.6	49.2	51.0	44.6
No response	1.2		3.1	1.2	

It can be seen that, in general, those involved in one of the discussion methods felt that this course necessitated more preparation than did those in the lecture sections or the television observation section. The results on these two questions were not completely consistent indicating, perhaps, that the wording of the questions confused the respondents to some extent. It was quite clear, however, that the television observation group felt least compelled to prepare for this course.

It is often said that students can acquire more facts from a course taught by the lecture method than one taught by the discussion method. The data from this questionnaire would seem to indicate that these students felt otherwise. Tables XVII and XVIII show that students in the small discussion sections felt that they learned most, while those in the television discussion section were only slightly less optimistic about their learning. Students in the television observation section were the most pessimistic among the groups. These differences were significant at the 5 per cent level of confidence. ($\chi^2 = 26.30$ and 26.60 , $df. = 8$.)

Significant also were the differences between these groups on the extent to which they felt motivated to greater thought by this course as

TABLE XVII

	TV Disc.	TV Obs.	Large Class Disc.	Lecture	Small Class Disc.
Do you feel that you learn:					
More facts	47.8%	18.5%	46.2%	38.6%	50.8%
About the same number	25.0	38.1	30.7	34.9	44.6
Fewer facts	26.3	44.4	20.0	24.5	4.6
in this course than in comparable <i>discussion</i> courses?					
No response	1.2		3.1	2.0	

TABLE XVIII

	TV Disc.	TV Obs.	Large Class Disc.	Lecture	Small Class Disc.
Do you feel that you learn:					
More facts	43.8%	25.9%	38.5%	42.6%	50.8%
About the same number	32.5	40.8	23.1	43.4	35.4
Fewer facts	22.5	33.3	33.8	13.6	12.3
in this course than in com- parable lecture courses?					
No response	1.2		4.6	.4	1.5

compared to other lecture and discussion courses. These data are shown in Tables XIX and XX. These differences were also significant at the 5 per cent level of confidence. ($\chi^2 = 67.08$ and 28.31 respectively, with 8 df.) All groups, except the small discussion group, apparently felt that discussion courses are more thought provoking than are lecture courses. On both comparisons, the television discussion group responded most favorably. On one comparison, the television observation group responded least favorably and on the other the small group discussion sections responded least favorably.

TABLE XIX

	TV Disc.	TV Obs.	Large Class Disc.	Lecture	Small Class Disc.
Do you feel that you are stimulated to:					
Greater thought	53.8%	22.2%	41.6%	26.1%	47.7%
About the same	37.5	22.2	49.2	38.2	41.6
Less thought	8.7	55.6	7.7	34.9	7.7
in this course than in com- parable discussion courses?					
No response			1.5	.8	3.0

TABLE XX

	TV Disc.	TV Obs.	Large Class Disc.	Lecture	Small Class Disc.
Do you feel that you are stimulated to:					
Greater thought	60.0%	37.1%	44.6%	36.9%	32.9%
About the same	27.5	33.3	30.7	44.2	38.5
Less thought	12.5	29.6	21.6	18.1	7.7
in this course than in com- parable lecture courses?					
No response			3.1	.8	1.5

Results on the instructor attitude criterion.

The subjective impressions of the instructors participating in more than one type of section in the experiment were solicited. Specifically, the opinions of the two instructors handling the various types of discussion sections, since these represented innovations in the "normal" teaching methods, were considered of interest. It is important to keep in mind that these reactions fall into two separate though related categories: (1) reactions to the discussion method, *per se*, and (2) reactions to the use of closed-circuit television.

With respect to the first, the most significant reaction of the instructor handling the small discussion sections in the classroom is perhaps best stated in his own words:

My first and outstanding impression of the small group discussion method for the beginning, introductory course in American Government was that it was somewhat less than adequate as a teaching device. A satisfactory discussion should be a many-sided, give-and-take performance between the instructor and the students. This, in turn, demands that the student have a somewhat better than rudimentary understanding and knowledge of the subject material in order that his responses and his questions be timely, challenging, and thoughtful. Therein lies the crux of the problem in a beginning course taught solely by the discussion technique. My experience indicated quite definitely that the average student in American Government was rather woefully lacking a basic understanding of the subject material. Far too often the response to a discussion question was merely an "I don't know" or a rather irrelevant foray into uncharted seas of shallow opinion and "off the cuff" impressions. Perhaps the fact that politics and political opinion are everybody's business seduced the student into believing that he need not tackle the subject material with vigor and effort, but instead could rely on a sort of passing knowledge to see him through the rigors of group discussion. In any event, the burden of discussion fell all too heavily on the few top students (who were no doubt immensely benefited by it), and a consequent tendency for the remaining students to retreat into apathy and hesitancy was evident.

In addition, this instructor mentioned that at times the lack of student response made the discussion method stultifying to him; often key questions raised did not receive more than fragmentary consideration through class discussion, and the result would be somewhat in the nature of a partial, unplanned lecture. It is perhaps emphasizing the obvious to say that discussion depends on well-prepared students, and, in general, only a minority of these undergraduates were academically equipped to make such discussion worth while.

If the reactions of this instructor can be generalized, it would appear that the discussion method for the type of course dealt with in this study should probably be confined to sections of superior students.

In comparing the television discussion section with the large group discussion section, the instructor here came to the conclusion that other variables were probably more important than the situational difference. During the first semester he felt that the large discussion section in the classroom was not at all worth while, at the same time being comparatively enthusiastic about the performance of the students in the television discussion section. His reaction during the second semester was almost the opposite. Whether this was because of the difference in students, his own possible improvement in directing large discussion groups combined with "creeping apathy" toward television, or some other possible explanation he was unable to say. In every instance, however, he would in general second the remarks concerning the discussion method itself made by the other instructor.

One of the greatest drawbacks of the television section was the feeling on the part of the instructor that he could never really establish rapport with the class. This to him is an important aspect of teaching, and his morale suffered from its absence. Comparing the two "best" sections of each method (i.e., television section of the first semester with the large discussion class of the second semester) he felt that only in the second did he establish rapport approximating that of the normal lecture-discussion method. The degree of importance of this reaction will, of course, depend upon the idiosyncracies of individual instructors; it seemed very important to this instructor.

One of the difficulties in promoting discussion using closed circuit television concerned the mechanical arrangements. Everything was done that could be done to speed up and simplify the mechanics. A viewing room proctor had only to press a switch and any student in the viewing room could speak and be heard by the instructor in the studio and other students in the studio and the viewing rooms. Sound was picked up by microphones hung permanently overhead. Nonetheless, the instructor felt that the time lag between a signal to the instructor that someone in one of the viewing rooms wished to recite and the actual recitation of that person inhibited a number of students from entering the discussion. Conceivably it might be more effective to reserve discussion to the studio group, confining communication from the viewing groups to specific questions or comments students in those groups wish to make.

With respect to those students who could only observe, reports reaching the instructor and the experimenter from the proctor in the room, from students in the group, and from observation, indicated increasing ennui during the course of the semester. Some students were even irritated because they did not have the opportunity to participate directly

in the discussion. The instructor indicated that during most of the semester he literally forgot the existence of this group.

Amount of participation criterion.

The three discussion methods were compared on the basis of the average number of student participations and the average number of different students participating each day. A sample of such participations was taken for a five day period during the first semester and a six day period during the second. For the comparison of the groups, in order that there would be an approximately equal number of students in each group that was compared, the participations in the three small group discussion sections were combined. During the first semester, there were no significant differences between the three groups on number of student participations each day. The groups did differ significantly on this criterion the second semester. Further analysis indicated that this was due to the fact that the television and large discussion groups had far more students than the total of the small discussion groups during the second semester. The television and large discussion groups did not differ from each other significantly, both simply had a significantly greater number of participations than did the total of the small discussion groups. During the first semester, when the three groups had approximately the same number of students, no significant differences were found.

The same results were obtained when the data were analyzed on the basis of the number of different persons participating in each group on the average day.

As noted above, part of the differences found for the second semester might be attributable to the differences in the number of subjects in each group. This would appear to be especially likely in the latter analysis based on the average number of different students participating each day. The larger groups would tend to have a proportionately greater number of students who like to participate. Another reasonable hypothesis for these results was the difference in the way in which each of the discussions was conducted. The instructor for the small group discussion sections did not attempt to get a great number of students to participate on any one day. He tried rather to assign certain students to be ready each day and concentrated on them during the discussion portions of the class. The instructor of the television and the large group discussion sections, on the other hand, did not single out special individuals for participation on any one day. He indicated, however, that he made a special effort to encourage discussion in the non-television group and tried to spread participation more. With the television group he made no similar effort. He gave no "pep talk" and seldom called on anyone except those students who raised their hands. Yet, participation in these latter two

groups was approximately equal. It should be noted also that no check was made of the amount of time each participant spoke. A five second comment and a five minute comment had equal weight in this analysis.

The second semester participation data from the large discussion group and the television discussion group were also analyzed separately. For the large discussion group, participations by students in the front and rear halves of the classroom were compared. It was felt that, though there was as much participation from this group as from the television group, the bulk of the participation might have been done by those sitting close to the instructor. (Seating in these sections was done alphabetically.) This hypothesis was not borne out by the data. There were no significant differences between the average daily participations or number of participants in the front and rear sections of the classroom. ($t = .91$ and $.37$ respectively.) As a matter of fact, participations from the rear half of the classroom averaged almost twice as many as those from the front half.

TABLE XXI

Group	<i>n</i>	Average no. of participations	Average no. of students participating
Studio	14	19.4	8
Viewing Group A	30	6.2	3.7
Viewing Group B	25	4.2	2.5
Total	69*	29.8	14.2

* This does not agree with the *n* cited in the description of the treatment groups earlier. There were three dropouts between the sampling period and the end of the semester.

A comparison of the participations from each viewing room and the studio were made for the television discussion group.

Table XXI shows the number in each group and the mean number of students participating each day and the mean number of participations.

During the first four days of the sampling period, student makeup of each part of the television discussion group remained unchanged. On the fifth day, students were rotated between the studio and the viewing rooms. During these first four days, it was found that 12 of the 14 students in the studio group participated at least once, 6 of the 30 students in Viewing Group A and 6 of the 25 in Viewing Group B participated at least once. Viewed in the light of the instructor's statement that he did not single out any individuals for participation, it would appear that it was much easier for those students in the studio to participate and/or they were much more motivated.

PART IV

Limitations of the Study

The primary limitation of this study is the same one found in many studies of educational methods, the possible inadequacy of the measuring instruments. In sections taught by different methods, especially if some of the methods involve discussion, the common ground on which it is possible to build examination items which measure achievement for all methods tends to be the textbook or other common sources of information. If there is an advantage to the discussion method of instruction, it probably does not lie in an increased rate of acquisition of knowledge which tends to be emphasized in regular course examinations. It was hoped that the attitude measure which is being developed by Professor Lane Davis would reflect something more than acquisition of knowledge. One is forced to conclude, however, that the instrument is still too gross to indicate possible differences between methods of instruction or that there are truly no differences. An examination of the instrument would lead one to accept the former hypothesis as at least a partial answer.

Another factor in this experiment which caused a heavy reliance on common sources of information for examination questions was that, as one instructor put it, "Political Science is a discipline essentially in a condition of anarchy; there is no consensus about what should go into any given course other than agreement on the broadest categories of subject matter. This means that given four different teachers of American Government, four different courses will be taught."

The conclusion which one is led to draw from all of the above is that the experiment here described might have shown primarily that these methods equally motivated students to study the text and other assignments.

The other factor which should be noted in the evaluation of these results is that the assignment of instructors to methods of instruction was not random and that, in general, instructor differences and method differences are confounded. The television instructor was assigned to this method because he had had experience on the medium. This same instructor was then assigned to the large group discussion method so that instructor differences between these methods, at least, could be minimized. Only one member of the political science department had been

assigned to teach three sections. For this reason, and because he had had experience in the discussion method of instruction, he was assigned to handle all of the small group discussion sections. This man was the junior faculty member in the group and the only instructor. Normally, the American Government course at the University is handled by all of the assistant professors on the political science staff.

PART V

Conclusions

Within the limitations of this study, as described, we must retain the hypothesis of no difference between these methods of instruction in their effect on acquisition of knowledge and attitude change. No significant differences were found on the amount of participation in the three discussion methods. The only significant differences found were in the area of student opinion. It was found that students prefer a course taught at least partially by the discussion method. Students receiving instruction by one of the discussion methods estimated the relative advantages of this course higher than did students in non-discussion sections. It also appeared that experience with receiving instruction by television tended to make students more favorable to this particular method of instruction.

Footnotes

1. A sample page of the syllabus can be found in the Appendix.
2. In those cases where one score was missing for a subject, he was arbitrarily assigned a score equal to the average for his group in that category for purposes of this evaluation.
3. E. F. Lindquist, *Design and Analysis of Experiments in Psychology and Education* (Boston: Houghton Mifflin Co., 1956), p. 269. Type I design.
4. *Ibid.*, pp. 121-123. Treatment by levels design.
5. Correlations between groups were found not to be homogeneous on this measure. Since this was the only test of homogeneity that exceeded the 5 per cent level of confidence, it is likely that this was a chance effect and that, in actuality, the groups are homogeneous.

Appendix

THE HISTORY OF THE CONSTITUTION

Text Assignment: Chapter 3, pp. 51-74.

U. S. Constitution, Art. 7 (p. 1078 of text)

Objectives of the Assignment:

1. To understand the meaning of constitutional government.
2. To understand the nature of the government of the Articles of Confederation and its weaknesses which made it unsatisfactory as a permanent constitution.
3. To become familiar with the events and motives that led to the movement to revise the Articles of Confederation.
4. To become familiar with the personnel of the Constitutional Convention, their backgrounds and interests, the areas of agreement and disagreement among the members of the Convention, and the procedures of the convention.
5. To understand the major compromises of the Constitution.
6. To become familiar with the major features of the campaign to ratify the Constitution.

Discussion questions:

1. Your text lists several features of the government under the Articles of Confederation which contributed to the failure of the Confederation government. What were these features of the government? How did each contribute to failure? Which do you think most important? Why?
2. Would you accept the verdict that the government of the Articles of Confederation was an unqualified failure? Why or why not?
3. What were the major causes for the movement for a revision of the Articles of Confederation? What political forces supported this movement and what particular revisions did they have in mind?
4. Identify three basic principles upon which the delegates to the Constitutional Convention of 1787 were agreed, and indicate (with specific references to the Constitution) how these principles were implemented.
5. In the United Nations, the smaller powers continually press for granting more strength to the organization. In the Confederation the smaller states were fearful of strengthening the national government. How do you account for this difference?

6. "The men who wrote the Constitution were truly remarkable politicians. They drafted a document which accommodated the conflicting interests of the most powerful pressure groups of their time and thus laid the foundation for a strong and stable national government." Do you agree? What were the major pressure groups at this time? What did they want? What were the issues on which differences existed?
7. Was the U.S. Constitution ratified in a legal and constitutional fashion? Explain. Should the extra-legal character of ratification cast doubt on the present validity of the Constitution? Why or why not?